

## Integrating Genetic and Transcriptomic Data to Identify Genes Underlying Obesity Risk

### Loci

#### AUTHORS:

Hanfei Xu, PhD<sup>1,\*</sup>, Shreyash Gupta, MS<sup>2</sup>, Ian Dinsmore, MS<sup>3</sup>, Abbey Kollu<sup>4</sup>, Anne Marie Cawley, MS<sup>5</sup>, Mohammad Y. Anwar, PhD<sup>6</sup>, Hung-Hsin Chen, PhD<sup>7,8</sup>, Lauren E. Petty, PhD<sup>8</sup>, Sudha Seshadri, PhD<sup>1,9,10</sup>, Misa Graff, PhD<sup>6</sup>, Piper Below, PhD<sup>8</sup>, Jennifer A. Brody, BA<sup>11</sup>, Geetha Chittoor, PhD<sup>2</sup>, Susan P. Fisher-Hoch, PhD<sup>12</sup>, Nancy L. Heard-Costa, PhD<sup>13,14</sup>, Daniel Levy, PhD<sup>15</sup>, Honghuang Lin, PhD<sup>16</sup>, Ruth JF. Loos, PhD<sup>17,18</sup>, Joseph B. McCormick, PhD<sup>12</sup>, Jerome I. Rotter, PhD<sup>19</sup>, Tooraj Mirshahi, PhD<sup>3</sup>, Christopher D. Still, DO<sup>20</sup>, Anita Destefano, PhD<sup>1,9</sup>, L. Adrienne Cupples, PhD<sup>1,‡</sup>, Karen L Mohlke, PhD<sup>21</sup>, Kari E. North, PhD<sup>6,†</sup>, Anne E. Justice, PhD<sup>2,†</sup>, Ching-Ti Liu, PhD<sup>1,†,\*</sup>

\* Corresponding authors

† Equally supervised this work

‡ Published posthumously

**Correspondence:** all correspondence should be addressed to [ctliu@bu.edu](mailto:ctliu@bu.edu) (CTL) and Hanfei Xu ([hfxu@bu.edu](mailto:hfxu@bu.edu))

#### AFFILIATIONS

<sup>1</sup>Department of Biostatistics, School of Public Health, Boston University, 801 Massachusetts Ave, Boston, MA, 02118, USA, <sup>2</sup>Department of Population Health Sciences, Geisinger, 100 N. Academy Ave., Danville, PA, 17822, USA, <sup>3</sup>Department of Genomic Health, Geisinger, 100 N.

Academy Ave., Danville, PA, 17822, USA, <sup>4</sup>Department of Psychology and Neuroscience, University of North Carolina, 235 E. Cameron Avenue, Chapel Hill, NC, 27599, USA, <sup>5</sup>Marsico Lung Institute, University of North Carolina, 125 Mason Farm Rd, Chapel Hill, NC, 27599, USA, <sup>6</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, 135 Dauer Drive, Chapel Hill, NC, 27599, USA, <sup>7</sup>Institute of Biomedical Sciences, Academia Sinica, No. 128, Section 2, Academia Rd., Taipei, Nangang District, 115201, Taiwan, <sup>8</sup>Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, 1161 21st Ave S, Nashville, TN, 37232, USA, <sup>9</sup>Department of Neurology, School of Medicine, Boston University, 85 East Concord Street, Boston, MA, 02118, USA, <sup>10</sup>Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, UT Health San Antonio, 8300 Floyd Curl Drive, San Antonio, TX, 78229, USA, <sup>11</sup>Department of Medicine, Cardiovascular Health Research Unit, University of Washington, 1730 Minor Ave, Seattle, WA, 98101, USA, <sup>12</sup>Department of Epidemiology, School of Public Health, UT Health Houston, Regional Academic Health Center, One West University Blvd, Brownsville, TX, 78520, USA, <sup>13</sup>Framingham Heart Study, 73 Mt Wayte Ave, Framingham, MA, 01702, USA, <sup>14</sup>Department of Neurology, Chobanian & Avedisian School of Medicine, Boston University, 72 E Concord St, Boston, MA, 02118, USA, <sup>15</sup>Population Sciences Branch, National Heart, Lung, and Blood Institute of the National Institutes of Health, 6701 Rockledge Drive, Bethesda, MD, 20892, USA, <sup>16</sup>Department of Medicine, University of Massachusetts Chan Medical School, 55 N Lake Ave, Worcester, MA, 01655, USA, <sup>17</sup>Charles Bronfman Institute for Personalized Medicine at Mount Sinai, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Pl, New York, NY, 10029, USA, <sup>18</sup>Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3A, 2200, Copenhagen, Denmark, <sup>19</sup>Department of Pediatrics, The Institute for

Translational Genomics and Population Sciences, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, 1124 West Carson Street, Torrance, CA, 90502, USA, <sup>20</sup>Center for Obesity and Metabolic Health, Geisinger, 100 N. Academy Ave., Danville, PA, 17822, USA, <sup>21</sup>Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill, NC, 27599, USA

## **AUTHOR CONTRIBUTIONS**

Participate in Project Concept and Design: LAC, KEN, AEJ, CTL. Participate in Parent Study Concept and Design: JBM, JIR, LAC, KEN. Participate in Phenotype Data Acquisition and/or QC: HX, ID, SPFH, RJFL, NHC, JBM, TM, CDS, LAC, AD, KEN, AEJ, CTL. Participate in Genotype or gene expression Data Acquisition and/or QC: HX, ID, MYA, HHC, LEP, MG, GC, HL, NHC, SPFH, AD, RJFL, JBM, TM, CDS, LAC, KEN, AEJ, CTL. Participate in Data Analysis and Interpretation: HX, SG, ID, MYA, RJFL, KLM, KEN, AEJ, CTL. Drafted the manuscript and revised according to co-author suggestions: HX, SG, SPFH, KEN, AEJ, CTL. All authors critically reviewed the manuscript, suggested revisions as needed, and approved the final version.

## **ACKNOWLEDGEMENTS**

### **Individual Acknowledgements:**

SG, ID, GC, CTL, KEN, AEJ, CDS, TM, KLM were funded in part by NIH R01 DK122503. MYA was funded by NIH NIDDK 3R01DK122503 – 02W1. MG was funded in part by NIH R01HL163262. JIR was funded in part by National Center for Minority Health Disparity (NCMHD), MD000170P20, McCormick. KEN was also funded in part by NIH R01HL142302, R01HL151152, R01 R01HD057194, R01HG010297, and R01HL143885.

### **Study Acknowledgements:**

The Framingham Heart Study (FHS): The FHS is funded by National Institutes of Health contract N01-HC-25195. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD. The analytical component of this project was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, and the Center for Information Technology, National Institutes of Health, Bethesda, MD. The visualization tools and data resources for this project were funded by the National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD.

The FHS acknowledges the support of contracts NO1-HC-25195, HHSN268201500001I and 75N92019D00031 from the National Heart, Lung and Blood Institute and grant supplement R01 HL092577-06S1 for this research. We also acknowledge the dedication of the FHS study participants without whom this research would not be possible. This manuscript does not necessarily reflect the opinions or views of the NHLBI, NIH or DHHS.

Cameron County Hispanic Community (CCHC) Cohort: We thank our cohort team including the CRU, data and lab staff of the CCHC team. We thank Valley Baptist Medical Center, Brownsville, for providing the space for our Clinical Research Unit. CCHC study and team members were funded by: National Institutes of Health, National Center for Minority Health Disparity, MD000170P20, UL1 TR000371, National Heart, Lung, and Blood Institute, 1R01HL142302, 2R01HL142302, National Center for Advancing Translational Sciences, (NCATS), Clinical and Translational Science Awards (CTSA), UL1 9/30/06-6/30/11, UL1TR0003716/27/12-/31/17, UL1TR0031677/1/19-6/30/2

MyCode Bariatric Surgery Cohort: We thank all the participants of the MyCode Community Health Initiative Study (MyCode) and the MyCode Research Team. We thank the members of the Geisinger-Regeneron DiscovEHR Collaboration who have been critical in the generation of the genetic and transcriptomic data used in this study. This study was funded in part by NIH HL142302, NIH 2R01HL142302, National Center for Advancing Translational Sciences, UL1 9/30/06-6/30/11, UL1TR0003716/27/12-/31/17, UL1TR0031677/1/19-(NCATS), Clinical and Translational Science Awards (CTSA).

## **CONFLICTS OF INTEREST**

None to report.

## ABSTRACT

Genome-wide association studies (GWAS) have identified numerous body mass index (BMI) loci. However, most underlying mechanisms from risk locus to BMI remain unknown. Leveraging omics data through integrative analyses could provide more comprehensive views of biological pathways on BMI. We analyzed genotype and blood gene expression data in up to 5,619 samples from the Framingham Heart Study (FHS). Using 3,992 single nucleotide polymorphisms (SNPs) at 97 BMI loci and 20,692 transcripts within 1 Mb, we performed separate association analyses of transcript with BMI and SNP with transcript ( $P_{\text{BMI}}$  and  $P_{\text{SNP}}$ , respectively) and then a correlated meta-analysis between the full summary data sets ( $P_{\text{META}}$ ). We identified transcripts that met Bonferroni-corrected significance for each omic, were more significant in the correlated meta-analysis than each omic, and were at least nominally associated with BMI in FHS data. Among 308 significant SNP-transcript-BMI associations, we identified seven genes (*NT5C2*, *GSTM3*, *SNAPC3*, *SPNS1*, *TMEM245*, *YPEL3*, and *ZNF646*) in five association regions. Using an independent sample of blood gene expression data, we validated results for *SNAPC3* and *YPEL3*. We tested for generalization of these associations in hypothalamus, nucleus accumbens, and liver and observed significant ( $P_{\text{META}} < 0.05$  &  $P_{\text{META}} < P_{\text{SNP}}$  &  $P_{\text{META}} < P_{\text{BMI}}$ ) results for *YPEL3* in nucleus accumbens and *NT5C2*, *SNAPC3*, *TMEM245*, *YPEL3*, and *ZNF646* in liver. The identified genes help link the genetic variation at obesity risk loci to biological mechanisms and health outcomes, thus translating GWAS findings to function.

## INTRODUCTION

Obesity is an enormous global public health burden.<sup>1</sup> Since obesity is a major risk factor for numerous health outcomes, including cardiometabolic diseases<sup>2</sup>, the rapid increase in the

global obesity burden requires immediate public health action and a better understanding of obesity pathogenicity to prevent it. Decades of research, including genome-wide association studies (GWAS), have demonstrated the fundamental role of genetic susceptibility in obesity risk.<sup>3-13</sup> Each GWAS-identified locus potentially provides novel biologic insight; yet identifying the functional variants, genes, and underlying pathways at these loci has limited translation for precision medicine.

A major barrier to precision medicine for obesity has been the identification of functional genes underlying GWAS findings. Of the thousands of genomic regions associated with obesity-related traits by GWAS, over 90% are in non-coding, potentially regulatory regions of the genome.<sup>14</sup> Previous work mapping body mass index (BMI)-related genes implicates the involvement of synaptic function and glutamate receptor signaling, which impinge on key hypothalamic circuits that respond to changes in feeding and fasting and are regulated by key obesity-related molecules such as *BDNF* and *MC4R*<sup>15-18</sup>. These pathways overlap with a proposed mechanism of action of topiramate, a component of one new FDA-approved weight-loss drug<sup>19,20</sup>. However, our understanding of the fundamental mechanisms underlying genetic risk for obesity is limited and controversial even for *FTO*, with the most prominent effects on BMI.<sup>21</sup>

Transcriptomics lie along pathways linking genetic susceptibility to obesity and is emerging as powerful disease biomarkers<sup>22-25</sup> that may provide targetable “mechanistic bridges” linking GWAS findings with obesity risk. Large-scale characterization and integration of OMICs have been challenging because the comprehensive collection of molecular data has, until very recently, either been unavailable or cost-prohibitive in the context of a single study. However, OMICs scans in the same individuals in which obesity-associated loci discoveries were made are now available<sup>24,26-30</sup>, thereby facilitating comprehensive and efficient integration with genetic data

to illuminate the underlying genes and mechanistic pathways of obesity-associated loci. Thus, studies that integrate GWAS with transcriptomics may lead to breakthroughs that reveal the genes contributing to obesity,<sup>31-33</sup> identify individuals or groups that could benefit from aggressive prevention or treatment, or the repurposing of therapeutics.<sup>34,35</sup>

In this study, we analyzed GWAS data and transcriptomic data generated in whole blood in 5,619 participants from the Framingham Heart Study (FHS) to identify potential causal genes through which known loci operate on obesity phenotypes (BMI). We used a correlated meta-analysis procedure to efficiently screen loci for potential candidate genes that are jointly associated with BMI and SNPs in linkage disequilibrium (LD) with established BMI-associated GWAS SNPs.

## **METHODS**

### ***Study sample***

We included participants from both the Offspring cohort and the third generation (Gen3) cohort of the FHS. The Offspring cohort of FHS began in 1971 and consisted of children of the Original cohort and spouses of these children.<sup>36,37</sup> Gen3 cohort comprised children from the offspring families enrolled in 2002.<sup>38</sup> The time intervals between clinical examinations for Offspring and Gen3 cohorts were approximately 4-6 years.

Since the timing of the blood sample taken for RNA collection was close to the eighth clinical examination (Exam 8) for the Offspring cohort and the second clinical examination (Exam 2) for the Gen3 cohort, our study was restricted to subjects with available blood sample, genotype data, and BMI information in either Exam 8 of the Offspring study or Exam 2 of the Gen3 study. All participants provided written informed consent. The Institutional Review Board of the Boston University Medical Campus approved the study (N=5,169).



All participants provided written informed consent. The Institutional Review Board of the Boston University Medical Campus approved the study.

### ***Data description***

FHS participants were genotyped using the Affymetrix GeneChip Human Mapping 500K Array Set and another Affymetrix 50K gene-centric array. The genotype imputation was performed using the Michigan Imputation Server with HRC reference panel release 1.1 April 2016 (HRC r1.1).

Fasting peripheral whole blood samples (2.5 ml) were collected from FHS participants at the eighth clinical examination (Exam 8) of the Offspring cohort and the second clinical examination (Exam 2) of the Gen3 cohort. The details of RNA collection and expression data cleaning have been previously described.<sup>39</sup> In our study, we used the expression data that have been adjusted using technical covariates and blood count<sup>40,41</sup>.

Height and weight were measured at Exam 8 of the Offspring cohort and Exam 2 of the Gen3 cohort. BMI was then calculated by weight (kg)/height(m)<sup>2</sup>.

### ***SNP-transcript association and transcript-BMI association***

We analyzed 3,992 SNPs that are in LD ( $r^2 > 0.8$ ) with 97 previously reported BMI variants from GIANT BMI GWAS paper (Locke et al. 2015) and the 1,408 transcripts with a start position within 1 Mb of these variants.

We performed two kinds of association modeling. The first was a SNP-transcript association model, with the transcript as the outcome, and the SNP genotype as the predictor, adjusting for covariates including age at expression data collection, sex, and cohort identifier. We

performed this first model for every SNP-transcript pair, using a linear mixed effects model to account for relatedness. The second model assessed the association between transcript and BMI, with expression of the transcript as the outcome, and BMI as the predictor, adjusting for age at expression data collection, sex, cohort identifier, and familial relatedness. We performed the second model for each transcript separately. In this manuscript, we will denote the p-value of the SNP from the first model as  $P_{SNP}$  and the p-value of BMI from the second model as  $P_{BMI}$ .

### *Correlated meta-analysis*

We used the correlated meta-analysis model of Province and Borecki<sup>42</sup> to account for the potential dependence between the SNP-transcript and transcript-BMI associations. This correlated meta-analysis model estimated the degree of correlation between SNP-transcript and transcript-BMI associations, and corrected for the inflation of type-I error that would be observed in a traditional meta-analysis (that assumes the two associations are statistically independent). Our model used a tetrachoric correlation, which was less sensitive to contamination from the alternative hypothesis than the Pearson correlation, thus preventing over-correction of the correlation.

In our analysis, for every SNP, we estimated the covariance matrix  $\Sigma$  between two association results ( $Z_{SNP} = \Phi^{-1}(P_{SNP})$  and  $Z_{BMI} = \Phi^{-1}(P_{BMI})$ ) using tetrachoric correlation, and then we calculated  $Z_{meta} = (Z_{SNP} + Z_{BMI}) \sim N(0, sum(\Sigma))$  and  $P_{meta} = 1 - \Phi(Z_{meta})$  for each SNP-transcript pair.

After performing the correlated meta-analysis, we further screened the results to identify transcripts that met Bonferroni-corrected significance for each omic and were more significant in the correlated meta-analysis than in each omic. Thus, we included five criteria:  $P_{meta} < P_{SNP}$ ,  $P_{meta} < P_{BMI}$ ,  $P_{SNP} < (0.05/1,408) = 3.6 \times 10^{-5}$ ,  $P_{BMI} < 3.6 \times 10^{-5}$ , and at least nominal association ( $P < .05$ ) between BMI and SNP in FHS. The first two criteria ensured that both the SNP-transcript and

transcript-BMI associations contributed to the meta-analysis. The third and fourth criteria guaranteed the Bonferroni-corrected significance of each association. The last criterion restricted the SNPs to those at least nominally associated with BMI in FHS data.

### ***Functional interrogation/regulatory annotation***

Regulatory variants are more likely to drive correlated signals of gene expression and SNP association than coding variants. To characterize candidate regulatory variants, we used chromatin marks and other epigenomic data that define regulatory elements or link regulatory elements to gene transcription start sites. We focused on data sets for liver, and component cell types, especially preadipocytes, adipocytes, and hepatocytes. We compared them to other tissues because tissue-restricted regulatory elements may be more likely to be relevant and functional. The resources we considered include accessible chromatin based on the assay for transposase-accessible chromatin (ATAC-seq) or DNase hypersensitivity from brain, blood, and liver; histone mark and transcription factor ChIP-seq and chromatin states from ENCODE<sup>43</sup> used for visual inspection and to assess variant overlap with potential candidate cis regulatory elements (cCREs). Additional resources used for variant annotation as described in Box 1 include GeneCards<sup>44</sup>, OMIM<sup>45</sup>, and GTEx<sup>46</sup>.

### ***Validation: Cameron County Hispanic Cohort (CCHC)***

The CCHC was established on the Texas-Mexico border in 2004.<sup>47</sup> This randomly ascertained community cohort currently comprises over 5000 people and is approximately 60% female. The mean age of the CCHC participants was 45.2 years and 61% were > age 40. All CCHC individuals were genotyped using the Illumina MEGAEX array.<sup>48</sup> All genotype data was quality

controlled (QC'ed) following standard protocols (i.e. exclude individuals and variants with high levels of missingness, extreme heterozygosity, sex mismatch, variants with low minor allele frequency (MAF <0.01), or those that deviate from Hardy-Weinberg equilibrium ( $p < 10^{-6}$ ), then imputed to TOPMed R2 panel.<sup>49</sup> The study was approved by the Committee for the Protection of Human Subjects (CPHS) at the University of Texas Health Sciences Center at Houston. All participants provided informed consent to be included in CCHC genomic studies.

RNA sequencing of 1,800 CCHC participants was conducted using stored whole blood with sufficient quantity and quality. Sample collection and transcriptome profiling were described in detail previously.<sup>50</sup> Briefly, pooled libraries were subjected to 150 bp paired-end sequencing according to the protocol (Illumina NovaSeq) at VANTAGE. Six blood cell types were predicted and scaled using the R package DeconCell, a method that quantifies cell types using expression of marker genes.<sup>51</sup> We performed fastp for quality control following standard protocols,<sup>52</sup> and the QC-passed reads were aligned to human genome reference (hg38) with STAR.<sup>53,54</sup> Samples with aligned reads <15M, alignment rate <40%, or assigned reads <15M were excluded.<sup>55</sup> DESeq2 was performed for library size normalization and gene-specific dispersion estimation.<sup>56</sup> PEER factor analysis was used to capture the unobserved confounders of transcriptome profiles.<sup>57</sup> We then implemented a negative binomial model in DESeq2 to identify BMI genes with covariate adjustment for sex, age, 10 PEER factors<sup>56</sup>, and filtered results using default thresholds.

We performed eQTL mapping using the GTEx v8 pipeline.<sup>46</sup> In brief, we selected 645 unrelated CCHC individuals with both genotyping and RNA-seq data available.<sup>58</sup> We aligned the RNA-seq reads to the human genome reference using STAR, and then quantified each gene using RNA-SeQC.<sup>53,59</sup> Read counts were normalized by trimmed mean of M values (TMM), and inverse variance normalization was performed.<sup>60</sup> We identified eQTLs in cis (within 1 Mb) for each gene

using FastQTL with adjustment for sex, RNA-seq batch, 5 genetic principal components (PCs), and 10 PEER factors.

### ***Generalization on liver tissue: MyCode Bariatric Surgery Cohort***

The MyCode™Community Health Initiative (MyCode) study is a healthcare-based population study in central and northeastern Pennsylvania with ~2 million patients<sup>61,62</sup>. All participants provided informed consent for the MyCode Study. This study was approved by the Geisinger Institutional Review Board. We leveraged existing transcriptomic profiling in the Geisinger Health System's (GHS) Bariatric Surgery Program (BSP) study to generalize observed joint associations from whole blood in FHS to liver tissue. All BSP participant data are linked to clinical and demographic data through Mycode's electronic health records. While the BSP study participants are all obese, there is significant variation across the study participants (**Supplementary Table 1**, N=2,224). Liver wedge biopsies were obtained intraoperatively during Roux-en-Y gastric bypass (RYGB) surgery<sup>63</sup>, as described previously<sup>64,65</sup>. Total RNA was isolated from the excised liver tissue using the RNeasy total RNA isolation kit (Qiagen, Valencia, CA). Standard library prep procedures were followed by sequencing to obtain raw reads for each sample. The sequencing reads were aligned to Ensembl Release 104 reference genotypes and duplicate reads were marked using STAR v2.7.0,<sup>53</sup> and then quality-controlled and gene expression quantified using and converted to transcripts per million (TPM) using RNASEQC v2.4.2<sup>59</sup>. TPM  $\geq 0.1$  in at least 20% of samples and  $\geq 6$  reads in at least 20% of samples. Raw TPM values were inverse normal transformed<sup>60</sup>. PEER factor analysis was used to capture the unobserved confounders of transcriptome profiles.<sup>57</sup> Association analyses were performed using FastQTL<sup>66</sup>,

adjusting for sex, age, self-identified race/ethnicity, the first three genomic PCs to control for ancestry, and 60 PEER factors.<sup>56</sup>

### ***Generalization on brain tissue***

Analyses of hypothalamus (N=131) and nucleus accumbens (N=198) were conducted on samples from three cohorts: the Framingham Heart Study (FHS), the Religious Orders Study (ROS), and the Rush Memory and Aging Project (MAP). We restricted our analysis to samples with RIN>3 and BRAAK score <= 4. Details of RNA sequencing of hypothalamus and nucleus accumbens and the transcript-BMI association analysis were described previously<sup>67</sup>. For the eQTL analysis, we used FastQTL and adjusted for covariates: 5 first genetic PCs, PEER factors according to the GTEx recommendations (15 PEER factors for hypothalamus and 30 PEER factors for nucleus accumbens), sex, age at death, cohort, and sequencing batch. We further performed meta-analysis using p-values of SNP-transcript and transcript-BMI associations via Fisher's method<sup>68</sup>, producing a meta-analyzed p-value.

## **RESULTS**

### ***Sample characteristics***

The characteristics of samples included in the discovery correlated meta-analysis, validation analysis, and generalization analyses are shown in **Supplementary Table 1**. The age distribution was similar for FHS whole blood, CCHC whole blood, and GHS liver analyses, with a mean ranging from 47 to 58, and the brain analyses had relatively older subjects with a mean age of 88. All study samples had a larger proportion of females compared to males. The GHS sample

had a relatively higher BMI compared to other study samples. FHS and GHS are dominantly European ancestry, while CCHC was 100% Hispanic/Latino.

### ***Correlated Meta-analysis***

**Figure 1** shows the general workflow of our entire study. The models and filtering criteria of each step have been included in the methods section. In the FHS analysis, we found 308 SNP-transcript-BMI associations corresponding to seven unique genes (*NT5C2*, *YPEL3*, *ZNF646*, *SPNS1*, *GSTM3*, *SNAPC3*, and *TMEM245*) potentially involved in transcriptional pathways from SNP to BMI (**Table 1**). 115 variants were involved in the SNP-transcript-BMI associations for *NT5C2*, including the reported BMI variant rs11191560<sup>69,70</sup>. *YPEL3*, *ZNF646* and *SPNS1* were in the same region (16p11.2), and we observed 10, 46 and 91 SNP-transcript-BMI associations for *YPEL3*, *ZNF646* and *SPNS1* respectively, including three reported BMI variants rs4787491<sup>28</sup>, rs9925964<sup>69-71</sup> and rs3888190<sup>28,69,71-74</sup>. At the *TMEM245* locus, we pinpointed the SNP-transcript-BMI association to the reported BMI variant rs6477694<sup>69-71,73-75</sup>. *GSTM3* was located at 1p13.3, with 4 SNP-transcript-BMI associations detected, including previously reported BMI signal rs17024393<sup>28,69,70,76,77</sup>. *SNAPC3*, located at 9p22.3, had 41 SNP-transcript-BMI associations identified. Separate results for suggestive ( $P_{\text{SNP}} < (0.05/1,408) = 3.6 \times 10^{-5}$ ,  $P_{\text{BMI}} < 3.6 \times 10^{-5}$ ) SNP-transcript and transcript-BMI signals are provided in **Supplementary Table 2** and **Supplementary Table 3**.

### ***Validation and Generalization to Other Tissues***

We tested for validation of the above seven genes using CCHC blood gene expression data. Among the identified 308 SNP-transcript-BMI associations, 37 SNP-transcript-BMI associations

corresponding to *SNAPC3* and 10 SNP-transcript-BMI associations corresponding to *YPEL3* remained significant ( $P_{\text{meta}} < 0.05$  &  $P_{\text{meta}} < P_{\text{SNP}} \& P_{\text{meta}} < P_{\text{BMI}}$ ) (**Supplementary Table 4**). Regional association plots for each gene show annotation information (**Figure 2** and **Supplementary Figures 1-5**). Of note, the top  $P_{\text{META}}$  SNP for *SNAPC3* and *YPEL3* are within or proximal to putative candidate cis-Regulatory Elements (cCREs) based on ENCODE<sup>43</sup> regulatory data on blood, brain, and liver tissues (**Figure 2**).

We also tested for generalization using gene expression in brain tissues. Hypothalamus tissue showed no significant SNP-transcript-BMI association. In contrast, the 10 SNP-transcript-BMI associations corresponding to *YPEL3* were significant in the generalization analysis on nucleus accumbens. Additionally, we were able to generalize signals in liver tissue for *NT5C2*, *SNAPC3*, *TMEM245*, *YPEL3*, and *ZNF646*, including 103, 40, 1, 10, and 15 SNP-transcript-BMI associations, respectively (**Supplementary Table 4**). While the direction of effect was consistent for both brain tissues, even for non-significant associations (**Table 2**, **Figure 3**, and **Supplementary Table 4**). The direction of effect was not always consistent across tissue types; however, consistency of direction of effect across various tissues may not be expected. Further work may be needed to clarify expectations of directional consistency across tissues with respect to BMI ~ Gene and SNP ~ associations.

### ***Biological interrogation***

Previous studies of gene function and bioinformatics characteristics (see Methods) of the significant genes highlight nearby signatures of gene regulation (**Box 1**, **Figure 2** and **Supplementary Figures 1-5**). Additionally, our literature search provides further details on potential roles for identified genes for obesity (**Box 1**). For example, *NT5C2* deletion was found



to be protective in mice fed a high fat diet (HFD) <sup>78</sup>. Other model organism studies have shown alterations in *YPEL3* results in altered obesity phenotypes. For example, *YPEL3* knockdown in *Drosophila melanogaster* resulted in significant changes in body fat percentage<sup>79</sup>. Additionally, previous studies integrating GWAS and eQTL data have shown *YPEL3* and *NT5C2* were both have pleiotropic effects leading to schizophrenia and cardiometabolic disease, schizophrenia and BMI for *YPEL3*, and schizophrenia, BMI, and coronary artery disease for *NT5C2*<sup>80</sup>. Box 1 shows strong support for *YPEL3* and *NT5C2* as likely candidate genes underlying the association with BMI in these two regions. However, existing knowledge that may offer a role for the other genes in the pathway to BMI is sparse.

## DISCUSSION

This study incorporated a correlated meta-analysis method to perform integrative analysis using genotype, gene expression, and phenotype (BMI) data. From the discovery analysis using the FHS whole blood data, we identified seven genes (*NT5C2*, *YPEL3*, *ZNF646*, *SPNS1*, *GSTM3*, *SNAPC3*, and *TMEM245*) that potentially lie along the pathway linking genetic variation to elevated BMI. Among those seven genes, *YPEL3* and *SNAPC3* associations were validated in whole blood in the CCHC study. In the analyses of tissues other than blood, *NT5C2*, *SNAPC3*, *TMEM245*, *YPEL3*, and *ZNF646* associations generalized in the liver tissue, and *YPEL3* in the nucleus accumbens.

*YPEL3* is located at 16p11.2, a gene dense region well-known for a microdeletion associated with neurocognitive developmental delay and predisposition to obesity<sup>81-83</sup>. Literature has reported that this region's deletion event is related to a highly-penetrant form of obesity<sup>84,85</sup>, and is age- and gender-dependent<sup>86,87</sup>. Within this region, *SH2B1* has received much attention as

the likely causal gene underlying the mosaic effects of the 16p11.2 deletion and is thought to regulate body weight and glucose metabolism<sup>88,89</sup>; and as a result, *YPEL3* has rarely been considered in previous studies. One of the previous studies that considered *YPEL3*<sup>80</sup> identified it as a pleiotropic gene jointly influencing BMI and risk of schizophrenia. In contrast, another study<sup>90</sup> asserted that the association between *YPEL3* and schizophrenia is due to its correlation with expression of *INO80E*, another possible candidate gene for BMI and risk of schizophrenia in the 16p11.2 region. Despite the controversial findings of *YPEL3* in the literature, several pieces of evidence support a role of *YPEL3* in BMI. First, the gene is highly expressed in whole blood and brain, similar to well-known BMI-related genes (**Box 1**). Also, *YPEL3* was the sole candidate gene in this region identified by the current analysis. Further, the blood expression results were validated in an independent study of Hispanic participants, and the results generalized to both brain and liver tissues. Combined, this evidence suggests that more attention is warranted on this gene in the future.

*NT5C2* is located at 10q24.32, which has been reported as a highlight locus of autism spectrum disorder, brain arterial diameters, and schizophrenia<sup>91-93</sup>. A previous study in zebrafish found *NT5C2* as a potential causal gene in this region for blood pressure<sup>94</sup>. Notably, variation in this gene is also associated with lower visceral and subcutaneous fat<sup>95</sup>, obesity, and the concurrence of obesity and depression<sup>96</sup> (**Box 1**). Further, animal studies of *NT5C2* knock-outs show changes in body weight gain, insulin resistance on high-fat diet, and white adipose tissue mass<sup>78,97</sup>. Kumar et al. found that rs11191548 decreased miRNA binding efficiency, which may explain the functional role of *NT5C2* influencing BMI<sup>98</sup>. Yet, our significant findings linking SNP variation to *NT5C2* gene expression with BMI in liver tissue is novel and a role for this gene in other tissues warrants further exploration.

While support for other genes identified herein is limited in the literature, *SNAPC3*, which validated in CCHC, and *TMEM245*, which generalized to liver tissue, have connections to obesity-related traits. For example, similar to both genes mentioned above, *SNAPC3* variants have also been associated with schizophrenia<sup>99</sup>. Also, DNA methylation in *SNAPC3* has been reported to mediate the association between breastfeeding and early-life growth trajectories<sup>100</sup>. The expression level of *TMEM245* has been associated with atrial fibrillation<sup>101</sup>, and schizophrenia-associated variants have been reported within this gene<sup>102</sup>.

In recent years, there has been growing interest in developing integrative approaches that utilize various OMICs data to uncover underlying biological mechanisms<sup>103-106</sup>. When individual-level data is available, combining multiple OMICs datasets to perform further analysis is preferred<sup>107,108</sup>. Yet, few integrative studies using summary-level data exist<sup>80,109</sup>, limiting cross study analyses. Thus, among all the integrative OMICs analyses, the correlation between OMICs is often ignored<sup>109,110</sup>. In our study, we leveraged the correlated meta-analysis framework proposed<sup>42</sup>, which is a robust approach to integrate “suspected” correlated SNP-transcript association and transcript-BMI association. This approach is useful for performing statistical integration and has been incorporated into many colocalization and polygenic pleiotropy detection methods<sup>111,112</sup>. By performing correlated meta-analysis using summary level data, we ensured the correlation between summary statistics of OMICs scans were considered.

There are some previous integrative studies on obesity-related phenotypes. Smemo et al<sup>21</sup>. found that obesity-associated variants within *FTO* were functionally connected with *IRX3* and *IRX5* expression. Voisin et al.<sup>113</sup> and Tang et al.<sup>114</sup> evaluated the association and the interaction between obesity-associated SNPs and DNA methylation changes. Kogelman et al.<sup>107</sup> detected co-expression patterns among eQTLs, integrated with protein data, and detected several obesity

candidate genes, such as *ENPPI*, *CTSL*, and *ABHD12B*. More recently, integrative analyses on multiple obesity and neuro-related phenotypes provided further gene lists that potentially affected relevant phenotypes jointly<sup>80,90</sup>. Also, a recent study colocalized splice junction quantitative trait loci (sQTLs) measured in subcutaneous adipose tissue with 24 BMI GWAS loci, including with *YPEL3*<sup>115</sup>, and another study has reported 162 BMI signals with a colocalized adipose eQTL.<sup>116</sup>

Compared to other integrative studies, our study has several strengths. To our knowledge, our study is the first one that takes the correlation between OMICs scans into the integrative analysis of BMI. We not only have a discovery study using whole blood samples from European ancestry, but validate these joint associations in an independent study of Hispanic participants, and generalize our findings to other tissues, including liver nucleus accumbens. Yet, our study has some limitations. First, the traditional meta-analysis instead of the correlated meta-analysis was used in the validation and generalization analyses due to data sparsity. Also, we only included two types of OMICs data in our analyses, genetics and gene expression data. However, these analyses gave us a comprehensive view of how our findings can be interpreted across ancestry and tissue type. And, our work offers a framework for future investigations incorporating additional of OMICs data, such as DNA methylation or protein data, as well as additional tissues, that can also be adopted for other traits of interest.

## CONCLUSION

Our study aimed to narrow in on causal genes that underly known obesity susceptibility loci. Specifically, we were interested in genetic variation that may be operating on variation in BMI through alterations in gene expression. Our integrative, multi-omics approach identified seven candidate genes within five genomic regions for BMI. Among these seven, we find the strongest

support for *YPEL3*, *NT5C2*, and *SNAPC3*, through validation across ancestries, generalization across BMI-relevant tissues, and/or existing literature with a connection to BMI-related traits or gene functions. This deep dive into the etiology of obesity risk loci gets us one step forward to connecting genetic variation to biological mechanisms and health outcomes, and thus translating GWAS findings to function so that obesity precision treatment and prevention can begin.

## REFERENCES

1. Collaboration, N.C.D.R.F. (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet* 390, 2627-2642. [10.1016/S0140-6736\(17\)32129-3](https://doi.org/10.1016/S0140-6736(17)32129-3).
2. Prospective Studies, C., Whitlock, G., Lewington, S., Sherliker, P., Clarke, R., Emberson, J., Halsey, J., Qizilbash, N., Collins, R., and Peto, R. (2009). Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet* 373, 1083-1096. [10.1016/S0140-6736\(09\)60318-4](https://doi.org/10.1016/S0140-6736(09)60318-4).
3. Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Magi, R., et al. (2010). Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature genetics* 42, 949-960. [10.1038/ng.685](https://doi.org/10.1038/ng.685).
4. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Magi, R., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* 42, 937-948. [10.1038/ng.686](https://doi.org/10.1038/ng.686).

5. Willer, C.J., Speliotes, E.K., Loos, R.J., Li, S., Lindgren, C.M., Heid, I.M., Berndt, S.I., Elliott, A.L., Jackson, A.U., Lamina, C., et al. (2009). Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature genetics* *41*, 25-34. 10.1038/ng.287.
6. Loos, R.J., Lindgren, C.M., Li, S., Wheeler, E., Zhao, J.H., Prokopenko, I., Inouye, M., Freathy, R.M., Attwood, A.P., Beckmann, J.S., et al. (2008). Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature genetics* *40*, 768-775. 10.1038/ng.140.
7. Lindgren, C.M., Heid, I.M., Randall, J.C., Lamina, C., Steinthorsdottir, V., Qi, L., Speliotes, E.K., Thorleifsson, G., Willer, C.J., Herrera, B.M., et al. (2009). Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS genetics* *5*, e1000508. 10.1371/journal.pgen.1000508.
8. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Magi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature* *518*, 187-196. 10.1038/nature14132.
9. Berndt, S.I., Gustafsson, S., Magi, R., Ganna, A., Wheeler, E., Feitosa, M.F., Justice, A.E., Monda, K.L., Croteau-Chonka, D.C., Day, F.R., et al. (2013). Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature genetics* *45*, 501-512. 10.1038/ng.2606.
10. Kilpelainen, T.O., Zillikens, M.C., Stancakova, A., Finucane, F.M., Ried, J.S., Langenberg, C., Zhang, W., Beckmann, J.S., Luan, J., Vandenput, L., et al. (2011). Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nature genetics* *43*, 753-760. 10.1038/ng.866.
11. Yang, J., Loos, R.J., Powell, J.E., Medland, S.E., Speliotes, E.K., Chasman, D.I., Rose, L.M., Thorleifsson, G., Steinthorsdottir, V., Magi, R., et al. (2012). FTO genotype is associated with phenotypic variability of body mass index. *Nature* *490*, 267-272. 10.1038/nature11401.

12. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* 28, 166-174. 10.1093/hmg/ddy327.
13. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., and Consortium, G. (2018). Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* 27, 3641-3649. 10.1093/hmg/ddy271.
14. Fall, T., and Ingelsson, E. (2014). Genome-wide association studies of obesity and metabolic syndrome. *Mol Cell Endocrinol* 382, 740-757. 10.1016/j.mce.2012.08.018.
15. Melo, C.V., Mele, M., Curcio, M., Comprido, D., Silva, C.G., and Duarte, C.B. (2013). BDNF regulates the expression and distribution of vesicular glutamate transporters in cultured hippocampal neurons. *PloS one* 8, e53793. 10.1371/journal.pone.0053793.
16. Shen, Y., Fu, W.Y., Cheng, E.Y., Fu, A.K., and Ip, N.Y. (2013). Melanocortin-4 receptor regulates hippocampal synaptic plasticity through a protein kinase A-dependent mechanism. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33, 464-472. 10.1523/JNEUROSCI.3282-12.2013.
17. Wu, Q., Clark, M.S., and Palmiter, R.D. (2012). Deciphering a neuronal circuit that mediates appetite. *Nature* 483, 594-597. 10.1038/nature10899.
18. Yang, Y., Atasoy, D., Su, H.H., and Sternson, S.M. (2011). Hunger states switch a flip-flop memory circuit via a synaptic AMPK-dependent positive feedback loop. *Cell* 146, 992-1003. 10.1016/j.cell.2011.07.039.
19. Gibbs, J.W., 3rd, Sombati, S., DeLorenzo, R.J., and Coulter, D.A. (2000). Cellular actions of topiramate: blockade of kainate-evoked inward currents in cultured hippocampal neurons. *Epilepsia* 41 Suppl 1, S10-16.

20. Poulsen, C.F., Simeone, T.A., Maar, T.E., Smith-Swintosky, V., White, H.S., and Schousboe, A. (2004). Modulation by topiramate of AMPA and kainate mediated calcium influx in cultured cerebral cortical, hippocampal and cerebellar neurons. *Neurochemical research* 29, 275-282.
21. Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gomez-Marin, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507, 371-375. [10.1038/nature13138](https://doi.org/10.1038/nature13138).
22. Lee, C.C., Watkins, S.M., Lorenzo, C., Wagenknecht, L.E., Il'yasova, D., Chen, Y.D., Haffner, S.M., and Hanley, A.J. (2016). Branched-Chain Amino Acids and Insulin Metabolism: The Insulin Resistance Atherosclerosis Study (IRAS). *Diabetes care* 39, 582-588. [10.2337/dc15-2284](https://doi.org/10.2337/dc15-2284).
23. Wang, T.J., Larson, M.G., Vasan, R.S., Cheng, S., Rhee, E.P., McCabe, E., Lewis, G.D., Fox, C.S., Jacques, P.F., Fernandez, C., et al. (2011). Metabolite Profiles and the Risk of Developing Diabetes. *Nature medicine* 17, 448-453. [10.1038/nm.2307](https://doi.org/10.1038/nm.2307).
24. Newgard, C.B. (2017). Metabolomics and Metabolic Diseases: Where Do We Stand? *Cell Metabolism* 25, 43-56. <https://doi.org/10.1016/j.cmet.2016.09.018>.
25. Kraus, W.E., Pieper, C.F., Huffman, K.M., Thompson, D.K., Kraus, V.B., Morey, M.C., Cohen, H.J., Ravussin, E., Redman, L.M., Bain, J.R., et al. (2016). Association of Plasma Small-Molecule Intermediate Metabolites With Age and Body Mass Index Across Six Diverse Study Populations. *The journals of gerontology. Series A, Biological sciences and medical sciences* 71, 1507-1513. [10.1093/gerona/glw031](https://doi.org/10.1093/gerona/glw031).
26. Shah, S.H., Kraus, W.E., and Newgard, C.B. (2012). Metabolomic profiling for the identification of novel biomarkers and mechanisms related to common cardiovascular diseases: form and function. *Circulation* 126, 1110-1120. [10.1161/circulationaha.111.060368](https://doi.org/10.1161/circulationaha.111.060368).
27. Lotta, L.A., Scott, R.A., Sharp, S.J., Burgess, S., Luan, J., Tillin, T., Schmidt, A.F., Imamura, F., Stewart, I.D., Perry, J.R., et al. (2016). Genetic Predisposition to an Impaired Metabolism of the



- Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. *PLoS medicine* *13*, e1002179. [10.1371/journal.pmed.1002179](https://doi.org/10.1371/journal.pmed.1002179).
28. Justice, A.E., Winkler, T.W., Feitosa, M.F., Graff, M., Fisher, V.A., Young, K., Barata, L., Deng, X., Czajkowski, J., Hadley, D., et al. (2017). Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun* *8*, 14977. [10.1038/ncomms14977](https://doi.org/10.1038/ncomms14977).
29. Mendelson, M.M., Marioni, R.E., Joehanes, R., Liu, C., Hedman, A.K., Aslibekyan, S., Demerath, E.W., Guan, W., Zhi, D., Yao, C., et al. (2017). Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. *PLoS medicine* *14*, e1002215. [10.1371/journal.pmed.1002215](https://doi.org/10.1371/journal.pmed.1002215).
30. van Leeuwen, E.M., Karssen, L.C., Deelen, J., Isaacs, A., Medina-Gomez, C., Mbarek, H., Kanterakis, A., Trompet, S., Postmus, I., Verweij, N., et al. (2015). Genome of The Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. *Nat Commun* *6*, 6065. [10.1038/ncomms7065](https://doi.org/10.1038/ncomms7065).
31. Howard, G., Moy, C.S., Howard, V.J., McClure, L.A., Kleindorfer, D.O., Kissela, B.M., Judd, S.E., Unverzagt, F.W., Soliman, E.Z., Safford, M.M., et al. (2016). Where to Focus Efforts to Reduce the Black-White Disparity in Stroke Mortality: Incidence Versus Case Fatality? *Stroke; a journal of cerebral circulation* *47*, 1893-1898. [10.1161/STROKEAHA.115.012631](https://doi.org/10.1161/STROKEAHA.115.012631).
32. Rodriguez, F., Olufade, T.O., Ramey, D.R., Friedman, H.S., Navaratnam, P., Heithoff, K., and Foody, J.M. (2016). Gender Disparities in Lipid-Lowering Therapy in Cardiovascular Disease: Insights from a Managed Care Population. *Journal of women's health*. [10.1089/jwh.2015.5282](https://doi.org/10.1089/jwh.2015.5282).
33. Dehlendorff, C., Andersen, K.K., and Olsen, T.S. (2015). Sex Disparities in Stroke: Women Have More Severe Strokes but Better Survival Than Men. *Journal of the American Heart Association* *4*. [10.1161/JAHA.115.001967](https://doi.org/10.1161/JAHA.115.001967).

34. Tardif, J.C., Rheume, E., Lemieux Perreault, L.P., Gregoire, J.C., Feroz Zada, Y., Asselin, G., Provost, S., Barhdadi, A., Rhains, D., L'Allier, P.L., et al. (2015). Pharmacogenomic determinants of the cardiovascular effects of dalcetrapib. *Circulation. Cardiovascular genetics* 8, 372-382. 10.1161/CIRCGENETICS.114.000663.
35. Moosavinasab, S., Patterson, J., Strouse, R., Rastegar-Mojarad, M., Regan, K., Payne, P.R., Huang, Y., and Lin, S.M. (2016). 'RE:fine drugs': an interactive dashboard to access drug repurposing opportunities. *Database : the journal of biological databases and curation* 2016. 10.1093/database/baw083.
36. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M., and Castelli, W.P. (1975). The Framingham Offspring Study. Design and preliminary data. *Prev Med* 4, 518-525.
37. DAWBER, T.R., MEADORS, G.F., and MOORE, F.E. (1951). Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health* 41, 279-281. 10.2105/ajph.41.3.279.
38. Splansky, G.L., Corey, D., Yang, Q., Atwood, L.D., Cupples, L.A., Benjamin, E.J., D'Agostino, R.B., Fox, C.S., Larson, M.G., Murabito, J.M., et al. (2007). The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* 165, 1328-1335. 10.1093/aje/kwm021.
39. Wang, L., Perez, J., Heard-Costa, N., Chu, A.Y., Joehanes, R., Munson, P.J., Levy, D., Fox, C.S., Cupples, L.A., and Liu, C.T. (2018). Integrating genetic, transcriptional, and biological information provides insights into obesity. *Int J Obes (Lond)*. 10.1038/s41366-018-0190-2.
40. Joehanes, R., Zhang, X., Huan, T., Yao, C., Ying, S.X., Nguyen, Q.T., Demirkale, C.Y., Feolo, M.L., Sharopova, N.R., Sturcke, A., et al. (2017). Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* 18, 16. 10.1186/s13059-016-1142-6.

41. Zhang, X., Joehanes, R., Chen, B.H., Huan, T., Ying, S., Munson, P.J., Johnson, A.D., Levy, D., and O'Donnell, C.J. (2015). Identification of common genetic variants controlling transcript isoform variation in human whole blood. *Nature genetics* 47, 345-352. 10.1038/ng.3220.
42. Province, M.A., and Borecki, I.B. (2013). A correlated meta-analysis strategy for data mining "OMIC" scans. *Pac Symp Biocomput*, 236-246.
43. ENCODE\_Project\_Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74. 10.1038/nature11247.
44. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* 54, 1 30 31-31 30 33. 10.1002/cpbi.5.
45. Amberger, J.S., Bocchini, C.A., Scott, A.F., and Hamosh, A. (2019). OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res* 47, D1038-D1043. 10.1093/nar/gky1151.
46. Consortium, G.T. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318-1330. 10.1126/science.aaz1776.
47. Chen, H.H., Petty, L.E., North, K.E., McCormick, J.B., Fisher-Hoch, S.P., Gamazon, E.R., and Below, J.E. (2022). Novel diabetes gene discovery through comprehensive characterization and integrative analysis of longitudinal gene expression changes. *Hum Mol Genet* 31, 3191-3205. 10.1093/hmg/ddac039.
48. Anwar, M.Y., Baldassari, A.R., Polikowsky, H.G., Sitlani, C.M., Highland, H.M., Chami, N., Chen, H.H., Graff, M., Howard, A.G., Jung, S.Y., et al. (2022). Genetic pleiotropy underpinning adiposity and inflammation in self-identified Hispanic/Latino populations. *BMC Med Genomics* 15, 192. 10.1186/s12920-022-01352-3.
49. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290-299. 10.1038/s41586-021-03205-y.

50. Sabotta, C.M., Kwan, S.Y., Petty, L.E., Below, J.E., Joon, A., Wei, P., Fisher-Hoch, S.P., McCormick, J.B., and Beretta, L. (2022). Genetic variants associated with circulating liver injury markers in Mexican Americans, a population at risk for non-alcoholic fatty liver disease. *Front Genet* *13*, 995488. [10.3389/fgene.2022.995488](https://doi.org/10.3389/fgene.2022.995488).
51. Aguirre-Gamboa, R., de Klein, N., di Tommaso, J., Claringbould, A., van der Wijst, M.G., de Vries, D., Brugge, H., Oelen, R., Vosa, U., Zorro, M.M., et al. (2020). Deconvolution of bulk blood eQTL effects into immune cell subpopulations. *BMC Bioinformatics* *21*, 243. [10.1186/s12859-020-03576-5](https://doi.org/10.1186/s12859-020-03576-5).
52. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* *34*, i884-i890. [10.1093/bioinformatics/bty560](https://doi.org/10.1093/bioinformatics/bty560).
53. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15-21. [10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635).
54. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* *30*, 923-930. [10.1093/bioinformatics/btt656](https://doi.org/10.1093/bioinformatics/btt656).
55. Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* *32*, 3047-3048. [10.1093/bioinformatics/btw354](https://doi.org/10.1093/bioinformatics/btw354).
56. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* *15*, 550. [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8).
57. Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* *7*, 500-507. [10.1038/nprot.2011.457](https://doi.org/10.1038/nprot.2011.457).

58. Staples, J., Nickerson, D.A., and Below, J.E. (2013). Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. *Genetic epidemiology* 37, 136-141. [10.1002/gepi.21684](https://doi.org/10.1002/gepi.21684).
59. DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M.D., Williams, C., Reich, M., Winckler, W., and Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 28, 1530-1532. [10.1093/bioinformatics/bts196](https://doi.org/10.1093/bioinformatics/bts196).
60. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11, R25. [10.1186/gb-2010-11-3-r25](https://doi.org/10.1186/gb-2010-11-3-r25).
61. Carey, D.J., Fetterolf, S.N., Davis, F.D., Faucett, W.A., Kirchner, H.L., Mirshahi, U., Murray, M.F., Smelser, D.T., Gerhard, G.S., and Ledbetter, D.H. (2016). The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med* 18, 906-913. [10.1038/gim.2015.187](https://doi.org/10.1038/gim.2015.187).
62. Dewey, F.E., Murray, M.F., Overton, J.D., Habegger, L., Leader, J.B., Fetterolf, S.N., O'Dushlaine, C., Van Hout, C.V., Staples, J., Gonzaga-Jauregui, C., et al. (2016). Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* 354. [10.1126/science.aaf6814](https://doi.org/10.1126/science.aaf6814).
63. Wood, G.C., Chu, X., Manney, C., Strodel, W., Petrick, A., Gabrielsen, J., Seiler, J., Carey, D., Argyropoulos, G., Benotti, P., et al. (2012). An electronic health record-enabled obesity database. *BMC Med Inform Decis Mak* 12, 45. [10.1186/1472-6947-12-45](https://doi.org/10.1186/1472-6947-12-45).
64. Leti, F., Malenica, I., Doshi, M., Courtright, A., Van Keuren-Jensen, K., Legendre, C., Still, C.D., Gerhard, G.S., and DiStefano, J.K. (2015). High-throughput sequencing reveals altered expression of hepatic microRNAs in nonalcoholic fatty liver disease-related fibrosis. *Transl Res* 166, 304-314. [10.1016/j.trsl.2015.04.014](https://doi.org/10.1016/j.trsl.2015.04.014).
65. Gerhard, G.S., Benotti, P., Wood, G.C., Chu, X., Argyropoulos, G., Petrick, A., Strodel, W.E., Gabrielsen, J.D., Ibele, A., Still, C.D., et al. (2014). Identification of novel clinical factors

- associated with hepatic fat accumulation in extreme obesity. *J Obes* 2014, 368210. 10.1155/2014/368210.
66. Ongen, H., Buil, A., Brown, A.A., Dermitzakis, E.T., and Delaneau, O. (2016). Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* 32, 1479-1485. 10.1093/bioinformatics/btv722.
67. Christian, W., Julie, A.S., Thor, D.S., Joli, B., Adam, L., Ann, M., Philip, L.D.J., David, A.B., Sudha, S., Richard, H.M., and Anita, L.D. (2022). RNA-sequencing of human post-mortem hypothalamus and nucleus accumbens identifies expression profiles associated with obesity. *bioRxiv*, 2022.2001.2008.473382. 10.1101/2022.01.08.473382.
68. Fisher, R.A. (1932). *Statistical methods for research workers*, 4th Edition (Oliver and Boyd).
69. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197-206. 10.1038/nature14177.
70. Hoffmann, T.J., Choquet, H., Yin, J., Banda, Y., Kvale, M.N., Glymour, M., Schaefer, C., Risch, N., and Jorgenson, E. (2018). A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci. *Genetics* 210, 499-515. 10.1534/genetics.118.301479.
71. Winkler, T.W., Justice, A.E., Graff, M., Barata, L., Feitosa, M.F., Chu, S., Czajkowski, J., Esko, T., Fall, T., Kilpelainen, T.O., et al. (2015). The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS genetics* 11, e1005378. 10.1371/journal.pgen.1005378.
72. Wood, A.R., Tyrrell, J., Beaumont, R., Jones, S.E., Tuke, M.A., Ruth, K.S., consortium, G., Yaghootkar, H., Freathy, R.M., Murray, A., et al. (2016). Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively. *Diabetologia* 59, 1214-1221. 10.1007/s00125-016-3908-5.

73. Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2017). Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nature genetics* *49*, 1458-1467. 10.1038/ng.3951.
74. Koskeridis, F., Evangelou, E., Said, S., Boyle, J.J., Elliott, P., Dehghan, A., and Tzoulaki, I. (2022). Pleiotropic genetic architecture and novel loci for C-reactive protein levels. *Nat Commun* *13*, 6939. 10.1038/s41467-022-34688-6.
75. Huang, J., Huffman, J.E., Huang, Y., Do Valle, I., Assimes, T.L., Raghavan, S., Voight, B.F., Liu, C., Barabasi, A.L., Huang, R.D.L., et al. (2022). Genomics and phenomics of body mass index reveals a complex disease network. *Nat Commun* *13*, 7973. 10.1038/s41467-022-35553-2.
76. Graff, M., Scott, R.A., Justice, A.E., Young, K.L., Feitosa, M.F., Barata, L., Winkler, T.W., Chu, A.Y., Mahajan, A., Hadley, D., et al. (2017). Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. *PLoS genetics* *13*, e1006528. 10.1371/journal.pgen.1006528.
77. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshihara, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M., et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. *Nature genetics* *53*, 1415-1424. 10.1038/s41588-021-00931-x.
78. Johanns, M., Kviklyte, S., Chuang, S.J., Corbeels, K., Jacobs, R., Herinckx, G., Vertommen, D., Schakman, O., Duparc, T., Cani, P.D., et al. (2019). Genetic deletion of soluble 5'-nucleotidase II reduces body weight gain and insulin resistance induced by a high-fat diet. *Mol Genet Metab* *126*, 377-387. 10.1016/j.ymgme.2019.01.017.
79. Baranski, T.J., Kraja, A.T., Fink, J.L., Feitosa, M., Lenzini, P.A., Borecki, I.B., Liu, C.T., Cupples, L.A., North, K.E., and Province, M.A. (2018). A high throughput, functional screen of human Body Mass Index GWAS loci using tissue-specific RNAi *Drosophila melanogaster* crosses. *PLoS genetics* *14*, e1007222. 10.1371/journal.pgen.1007222.
80. Liu, H., Sun, Y., Zhang, X., Li, S., Hu, D., Xiao, L., Chen, Y., He, L., and Wang, D.W. (2020). Integrated Analysis of Summary Statistics to Identify Pleiotropic Genes and Pathways for the

- Comorbidity of Schizophrenia and Cardiometabolic Disease. *Front Psychiatry* *11*, 256. 10.3389/fpsy.2020.00256.
81. Christian, S.L., Brune, C.W., Sudi, J., Kumar, R.A., Liu, S., Karamohamed, S., Badner, J.A., Matsui, S., Conroy, J., McQuaid, D., et al. (2008). Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol Psychiatry* *63*, 1111-1117. 10.1016/j.biopsych.2008.01.009.
82. Stoppel, L.J., Kazdoba, T.M., Schaffler, M.D., Preza, A.R., Heynen, A., Crawley, J.N., and Bear, M.F. (2018). R-Baclofen Reverses Cognitive Deficits and Improves Social Interactions in Two Lines of 16p11.2 Deletion Mice. *Neuropsychopharmacology* *43*, 513-524. 10.1038/npp.2017.236.
83. Bochukova, E.G., Huang, N., Keogh, J., Henning, E., Purmann, C., Blaszczyk, K., Saeed, S., Hamilton-Shield, J., Clayton-Smith, J., O'Rahilly, S., et al. (2010). Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* *463*, 666-670. 10.1038/nature08689.
84. Walters, R.G., Jacquemont, S., Valsesia, A., de Smith, A.J., Martinet, D., Andersson, J., Falchi, M., Chen, F., Andrieux, J., Lobbens, S., et al. (2010). A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* *463*, 671-675. 10.1038/nature08727.
85. Maillard, A.M., Hippolyte, L., Rodriguez-Herreros, B., Chawner, S.J., Dremmel, D., Aguera, Z., Fagundo, A.B., Pain, A., Martin-Brevet, S., Hilbert, A., et al. (2016). 16p11.2 Locus modulates response to satiety before the onset of obesity. *Int J Obes (Lond)* *40*, 870-876. 10.1038/ijo.2015.247.
86. Yu, Y., Zhu, H., Miller, D.T., Gusella, J.F., Platt, O.S., Wu, B.L., Shen, Y., and Children's Hospital Boston Genotype Phenotype Study, G. (2011). Age- and gender-dependent obesity in individuals with 16p11.2 deletion. *J Genet Genomics* *38*, 403-409. 10.1016/j.jgg.2011.08.003.
87. Perrone, L., Marzuillo, P., Grandone, A., and del Giudice, E.M. (2010). Chromosome 16p11.2 deletions: another piece in the genetic puzzle of childhood obesity. *Ital J Pediatr* *36*, 43. 10.1186/1824-7288-36-43.
88. Rui, L. (2014). SH2B1 regulation of energy balance, body weight, and glucose metabolism. *World J Diabetes* *5*, 511-526. 10.4239/wjd.v5.i4.511.



89. Jiang, L., Su, H., Wu, X., Shen, H., Kim, M.H., Li, Y., Myers, M.G., Jr., Owyang, C., and Rui, L. (2020). Leptin receptor-expressing neuron *Sh2b1* supports sympathetic nervous system and protects against obesity and metabolic disease. *Nat Commun* *11*, 1517. [10.1038/s41467-020-15328-3](https://doi.org/10.1038/s41467-020-15328-3).
90. Vysotskiy, M., Zhong, X., Miller-Fleming, T.W., Zhou, D., Autism Working Group of the Psychiatric Genomics, C., Bipolar Disorder Working Group of the Psychiatric Genomics, C., Schizophrenia Working Group of the Psychiatric Genomics, C., Cox, N.J., and Weiss, L.A. (2021). Integration of genetic, transcriptomic, and clinical data provides insight into 16p11.2 and 22q11.2 CNV genes. *Genome Med* *13*, 172. [10.1186/s13073-021-00972-1](https://doi.org/10.1186/s13073-021-00972-1).
91. Autism Spectrum Disorders Working Group of The Psychiatric Genomics, C. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism* *8*, 21. [10.1186/s13229-017-0137-9](https://doi.org/10.1186/s13229-017-0137-9).
92. Liu, M., Khasiyev, F., Sariya, S., Spagnolo-Allende, A., Sanchez, D.L., Andrews, H., Yang, Q., Beiser, A., Qiao, Y., Thomas, E.A., et al. (2023). Chromosome 10q24.32 Variants Associate with Brain Arterial Diameters in Diverse Populations: A Genome-Wide Association Study. *medRxiv*. [10.1101/2023.01.31.23285251](https://doi.org/10.1101/2023.01.31.23285251).
93. Wang, J., Liu, J., Li, S., Li, X., Yang, J., Dang, X., Mu, C., Li, Y., Li, K., Li, J., et al. (2023). Genetic regulatory and biological implications of the 10q24.32 schizophrenia risk locus. *Brain* *146*, 1403-1419. [10.1093/brain/awac352](https://doi.org/10.1093/brain/awac352).
94. Vishnolia, K.K., Hoene, C., Tarhbalouti, K., Revenstorff, J., Aherrahrou, Z., and Erdmann, J. (2020). Studies in Zebrafish Demonstrate That *CNNM2* and *NT5C2* Are Most Likely the Causal Genes at the Blood Pressure-Associated Locus on Human Chromosome 10q24.32. *Front Cardiovasc Med* *7*, 135. [10.3389/fcvm.2020.00135](https://doi.org/10.3389/fcvm.2020.00135).
95. Hotta, K., Kitamoto, A., Kitamoto, T., Mizusawa, S., Teranishi, H., Matsuo, T., Nakata, Y., Hyogo, H., Ochi, H., Nakamura, T., et al. (2012). Genetic variations in the *CYP17A1* and *NT5C2* genes

- are associated with a reduction in visceral and subcutaneous fat areas in Japanese women. *J Hum Genet* 57, 46-51. 10.1038/jhg.2011.127.
96. Samaan, Z., Lee, Y.K., Gerstein, H.C., Engert, J.C., Bosch, J., Mohan, V., Diaz, R., Yusuf, S., Anand, S.S., Meyre, D., and Epi, D.G.I. (2015). Obesity genes and risk of major depressive disorder in a multiethnic population: a cross-sectional study. *J Clin Psychiatry* 76, e1611-1618. 10.4088/JCP.14m09720.
97. Camici, M., Garcia-Gil, M., Allegrini, S., Pesi, R., and Tozzi, M.G. (2020). Evidence for a Cross-Talk Between Cytosolic 5'-Nucleotidases and AMP-Activated Protein Kinase. *Front Pharmacol* 11, 609849. 10.3389/fphar.2020.609849.
98. Kumar, P., Traurig, M., and Baier, L.J. (2020). Identification and functional validation of genetic variants in potential miRNA target sites of established BMI genes. *Int J Obes (Lond)* 44, 1191-1195. 10.1038/s41366-019-0488-8.
99. Goes, F.S., McGrath, J., Avramopoulos, D., Wolyniec, P., Pirooznia, M., Ruczinski, I., Nestadt, G., Kenny, E.E., Vacic, V., Peters, I., et al. (2015). Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am J Med Genet B Neuropsychiatr Genet* 168, 649-659. 10.1002/ajmg.b.32349.
100. Briollais, L., Rustand, D., Allard, C., Wu, Y., Xu, J., Rajan, S.G., Hivert, M.F., Doyon, M., Bouchard, L., McGowan, P.O., et al. (2021). DNA methylation mediates the association between breastfeeding and early-life growth trajectories. *Clin Epigenetics* 13, 231. 10.1186/s13148-021-01209-z.
101. Zhang, P.P., Sun, J., and Li, W. (2020). Genome-wide profiling reveals atrial fibrillation-related circular RNAs in atrial appendages. *Gene* 728, 144286. 10.1016/j.gene.2019.144286.
102. Xu, C., Aragam, N., Li, X., Villa, E.C., Wang, L., Briones, D., Petty, L., Posada, Y., Arana, T.B., Cruz, G., et al. (2013). BCL9 and C9orf5 are associated with negative symptoms in schizophrenia: meta-analysis of two genome-wide association studies. *PloS one* 8, e51674. 10.1371/journal.pone.0051674.

103. Lopez de Maturana, E., Alonso, L., Alarcon, P., Martin-Antoniano, I.A., Pineda, S., Piorno, L., Calle, M.L., and Malats, N. (2019). Challenges in the Integration of Omics and Non-Omics Data. *Genes (Basel)* *10*. 10.3390/genes10030238.
104. Buescher, J.M., and Driggers, E.M. (2016). Integration of omics: more than the sum of its parts. *Cancer Metab* *4*, 4. 10.1186/s40170-016-0143-y.
105. Graw, S., Chappell, K., Washam, C.L., Gies, A., Bird, J., Robeson, M.S., 2nd, and Byrum, S.D. (2021). Multi-omics data integration considerations and study design for biological systems and disease. *Mol Omics* *17*, 170-185. 10.1039/d0mo00041h.
106. Albuquerque, D., Nobrega, C., Manco, L., and Padez, C. (2017). The contribution of genetics and environment to obesity. *Br Med Bull* *123*, 159-173. 10.1093/bmb/ldx022.
107. Kogelman, L.J., Zhernakova, D.V., Westra, H.J., Cirera, S., Fredholm, M., Franke, L., and Kadarmideen, H.N. (2015). An integrative systems genetics approach reveals potential causal genes and pathways related to obesity. *Genome Med* *7*, 105. 10.1186/s13073-015-0229-0.
108. Zhang, X., Zhou, Z., Xu, H., and Liu, C.T. (2022). Integrative clustering methods for multi-omics data. *Wiley Interdiscip Rev Comput Stat* *14*. 10.1002/wics.1553.
109. Wang, B., Lunetta, K.L., Dupuis, J., Lubitz, S.A., Trinquart, L., Yao, L., Ellinor, P.T., Benjamin, E.J., and Lin, H. (2020). Integrative Omics Approach to Identifying Genes Associated With Atrial Fibrillation. *Circ Res* *126*, 350-360. 10.1161/CIRCRESAHA.119.315179.
110. Clark, C., Dayon, L., Masoodi, M., Bowman, G.L., and Popp, J. (2021). An integrative multi-omics approach reveals new central nervous system pathway alterations in Alzheimer's disease. *Alzheimers Res Ther* *13*, 71. 10.1186/s13195-021-00814-7.
111. Wang, F., Panjwani, N., Wang, C., Sun, L., and Strug, L.J. (2022). A flexible summary statistics-based colocalization method with application to the mucin cystic fibrosis lung disease modifier locus. *Am J Hum Genet* *109*, 253-269. 10.1016/j.ajhg.2021.12.012.

112. Foley, C.N., Staley, J.R., Breen, P.G., Sun, B.B., Kirk, P.D.W., Burgess, S., and Howson, J.M.M. (2021). A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun* 12, 764. [10.1038/s41467-020-20885-8](https://doi.org/10.1038/s41467-020-20885-8).
113. Voisin, S., Almen, M.S., Zheleznyakova, G.Y., Lundberg, L., Zarei, S., Castillo, S., Eriksson, F.E., Nilsson, E.K., Bluher, M., Bottcher, Y., et al. (2015). Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers. *Genome Med* 7, 103. [10.1186/s13073-015-0225-4](https://doi.org/10.1186/s13073-015-0225-4).
114. Tang, Y., Jin, B., Zhou, L., and Lu, W. (2017). MeQTL analysis of childhood obesity links epigenetics with a risk SNP rs17782313 near MC4R from meta-analysis. *Oncotarget* 8, 2800-2806. [10.18632/oncotarget.13742](https://doi.org/10.18632/oncotarget.13742).
115. Brotman, S.M., Raulerson, C.K., Vadlamudi, S., Currin, K.W., Shen, Q., Parsons, V.A., Iyengar, A.K., Roman, T.S., Furey, T.S., Kuusisto, J., et al. (2022). Subcutaneous adipose tissue splice quantitative trait loci reveal differences in isoform usage associated with cardiometabolic traits. *Am J Hum Genet* 109, 66-80. [10.1016/j.ajhg.2021.11.019](https://doi.org/10.1016/j.ajhg.2021.11.019).
116. Brotman, S.M., El-Sayed Moustafa, J.S., Guan, L., Broadaway, K.A., Wang, D., Jackson, A.U., Welch, R., Currin, K.W., Tomlinson, M., Vadlamudi, S., et al. (2023). Adipose tissue eQTL meta-analysis reveals the contribution of allelic heterogeneity to gene expression regulation and cardiometabolic traits. *bioRxiv*. [10.1101/2023.10.26.563798](https://doi.org/10.1101/2023.10.26.563798).
117. Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2020). Ensembl 2020. *Nucleic Acids Res* 48, D682-D688. [10.1093/nar/gkz966](https://doi.org/10.1093/nar/gkz966).
118. Kulkarni, S.S., Karlsson, H.K., Szekeres, F., Chibalin, A.V., Krook, A., and Zierath, J.R. (2011). Suppression of 5'-nucleotidase enzymes promotes AMP-activated protein kinase (AMPK) phosphorylation and metabolism in human and mouse skeletal muscle. *J Biol Chem* 286, 34567-34574. [10.1074/jbc.M111.268292](https://doi.org/10.1074/jbc.M111.268292).

119. Park, J.M., Park, D.H., Song, Y., Kim, J.O., Choi, J.E., Kwon, Y.J., Kim, S.J., Lee, J.W., and Hong, K.W. (2021). Understanding the genetic architecture of the metabolically unhealthy normal weight and metabolically healthy obese phenotypes in a Korean population. *Sci Rep* *11*, 2279. 10.1038/s41598-021-81940-y.
120. Wen, W., Zheng, W., Okada, Y., Takeuchi, F., Tabara, Y., Hwang, J.Y., Dorajoo, R., Li, H., Tsai, F.J., Yang, X., et al. (2014). Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Hum Mol Genet* *23*, 5492-5504. 10.1093/hmg/ddu248.
121. Blanco-Sánchez, B., Clément, A., Stednitz, S.J., Kyle, J., Peirce, J.L., McFadden, M., Wegner, J., Phillips, J.B., Macnamara, E., Huang, Y., et al. (2020). yippee like 3 (ypel3) is a novel gene required for myelinating and perineurial glia development. *PLoS genetics* *16*, e1008841. 10.1371/journal.pgen.1008841.
122. Li, C.Y., Ou, R.W., Chen, Y.P., Gu, X.J., Wei, Q.Q., Cao, B., Zhang, L.Y., Hou, Y.B., Liu, K.C., Chen, X.P., et al. (2021). Genetic Analysis of ZNF Protein Family Members for Early-Onset Parkinson's Disease in Chinese Population. *Mol Neurobiol* *58*, 3435-3442. 10.1007/s12035-021-02354-5.
123. Langmyhr, M., Henriksen, S.P., Cappelletti, C., van de Berg, W.D.J., Pihlstrøm, L., and Toft, M. (2021). Allele-specific expression of Parkinson's disease susceptibility genes in human brain. *Sci Rep* *11*, 504. 10.1038/s41598-020-79990-9.
124. Perry, J.R., Day, F., Elks, C.E., Sulem, P., Thompson, D.J., Ferreira, T., He, C., Chasman, D.I., Esko, T., Thorleifsson, G., et al. (2014). Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* *514*, 92-97. 10.1038/nature13545.
125. Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghziyan, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics* *50*, 1112-1121. 10.1038/s41588-018-0147-3.

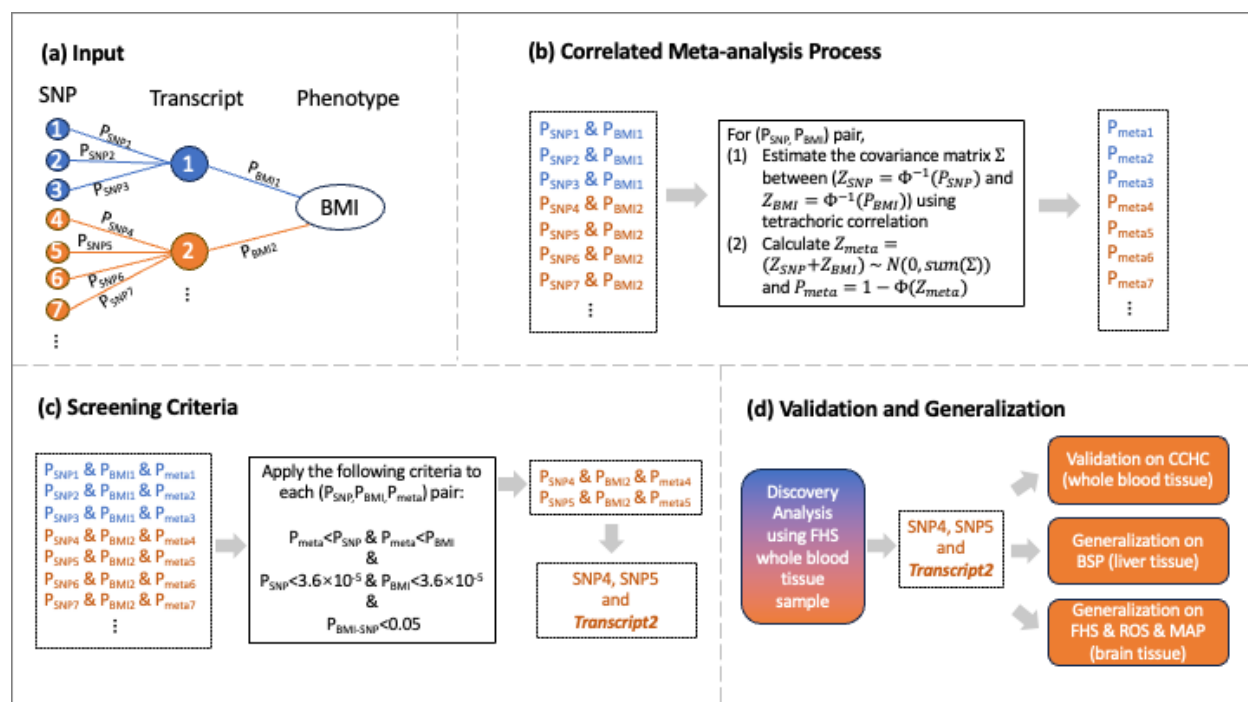
126. Huan, T., Liu, C., Joehanes, R., Zhang, X., Chen, B.H., Johnson, A.D., Yao, C., Courchesne, P., O'Donnell, C.J., Munson, P.J., and Levy, D. (2015). A systematic heritability analysis of the human whole blood transcriptome. *Hum Genet* *134*, 343-358. [10.1007/s00439-014-1524-3](https://doi.org/10.1007/s00439-014-1524-3).
127. Eldakhakhny, B.M., Al Sadoun, H., Choudhry, H., and Mobashir, M. (2021). In-Silico Study of Immune System Associated Genes in Case of Type-2 Diabetes With Insulin Action and Resistance, and/or Obesity. *Front Endocrinol (Lausanne)* *12*, 641888. [10.3389/fendo.2021.641888](https://doi.org/10.3389/fendo.2021.641888).
128. Zhu, Z., Guo, Y., Shi, H., Liu, C.L., Panganiban, R.A., Chung, W., O'Connor, L.J., Himes, B.E., Gazal, S., Hasegawa, K., et al. (2020). Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J Allergy Clin Immunol* *145*, 537-549. [10.1016/j.jaci.2019.09.035](https://doi.org/10.1016/j.jaci.2019.09.035).
129. Ferreira, M.A.R., Vonk, J.M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J.D., Helmer, Q., Tillander, A., Ullemar, V., Lu, Y., et al. (2019). Eleven loci with new reproducible genetic associations with allergic disease risk. *J Allergy Clin Immunol* *143*, 691-699. [10.1016/j.jaci.2018.03.012](https://doi.org/10.1016/j.jaci.2018.03.012).
130. Rao, S., Baranova, A., Yao, Y., Wang, J., and Zhang, F. (2022). Genetic Relationships between Attention-Deficit/Hyperactivity Disorder, Autism Spectrum Disorder, and Intelligence. *Neuropsychobiology* *81*, 484-496. [10.1159/000525411](https://doi.org/10.1159/000525411).
131. Pass, G.J., Becker, W., Kluge, R., Linnartz, K., Plum, L., Giesen, K., and Joost, H.G. (2002). Effect of hyperinsulinemia and type 2 diabetes-like hyperglycemia on expression of hepatic cytochrome p450 and glutathione s-transferase isoforms in a New Zealand obese-derived mouse backcross population. *J Pharmacol Exp Ther* *302*, 442-450. [10.1124/jpet.102.033553](https://doi.org/10.1124/jpet.102.033553).
132. Delles, C., Padmanabhan, S., Lee, W.K., Miller, W.H., McBride, M.W., McClure, J.D., Brain, N.J., Wallace, C., Marçano, A.C., Schmieder, R.E., et al. (2008). Glutathione S-transferase variants and hypertension. *J Hypertens* *26*, 1343-1352. [10.1097/HJH.0b013e3282fe1d67](https://doi.org/10.1097/HJH.0b013e3282fe1d67).
133. Cortón, M., Botella-Carretero, J.I., López, J.A., Camafeita, E., San Millán, J.L., Escobar-Morreale, H.F., and Peral, B. (2008). Proteomic analysis of human omental adipose tissue in the polycystic

ovary syndrome using two-dimensional difference gel electrophoresis and mass spectrometry. Hum

Reprod 23, 651-661. 10.1093/humrep/dem380.

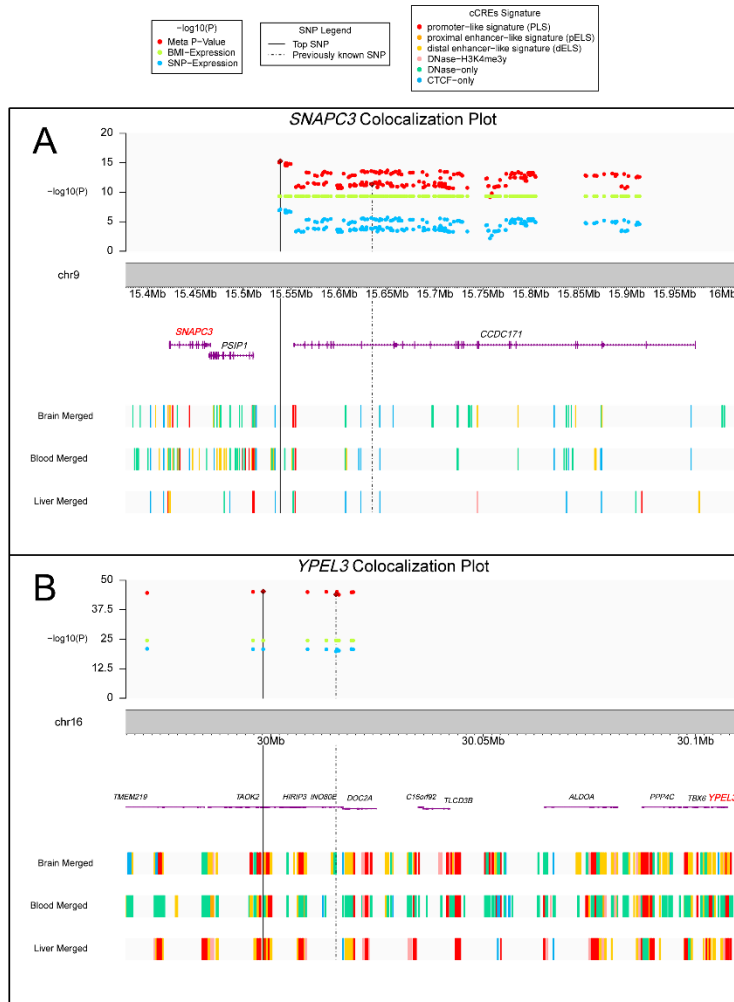
## FIGURES

**Figure 1. General workflow of the study design.** A) Step 1 included single omics associations for SNP to gene expression ( $P_{SNP}$ ) and gene expression to BMI ( $P_{BMI}$ ). B) Step 2 included the correlated meta-analysis to account for the interdependence between  $P_{SNP}$  and  $P_{BMI}$ . C) Identifying all SNP – Gene – BMI combinations that met our filtering criteria, which included correlated meta-analysis results that are more significant than individual omics associations. D) All significant SNP – Gene – BMI combinations were followed by validation in blood, liver, and brain tissues.

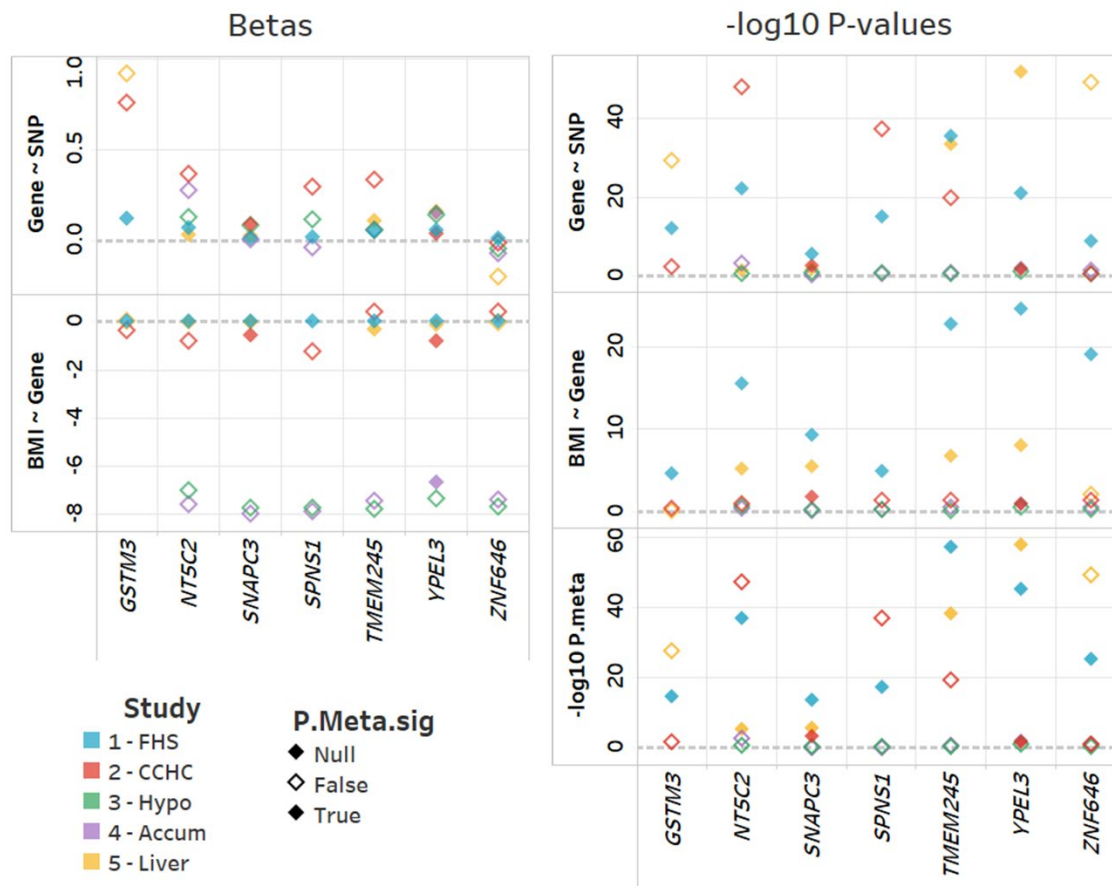




**Figure 2. Regional association plot** including association results for the discovery sample (Framingham Heart Study) for SNP with gene expression (blue), gene expression with BMI (green), and the correlated meta-analysis for SNP ~ gene expression ~ BMI (red). Annotation for potential candidate cis-regulatory elements from ENCODE are included for each reported SNP in the region. A. *SNAPC3*, B. *YPEL3*.



**Figure 3. Summary of validation and generalization for most significant SNP in discovery correlated meta-analysis.** Results are provided for discovery sample (FHS, blue), validation in blood (CCHC, red), and generalization to hypothalamus (Hypo, green), nucleus accumbens (Accum, purple), and liver (yellow) tissues. We provide individual effect estimates and P-values for each ‘omic and meta-analysis. Filled diamonds indicate significant associations in the meta-analysis (Note: FHS is noted as NULL, as all are significant).



## BOX

### Box 1. Seven genes identified as significant in the discovery analysis

***NT5C2. 5'-nucleotidase, cytosolic II*** is a protein-coding gene that may maintain internal composition of nucleotides. It hydrolyzes IMP (inosine monophosphate) and other purines (Genecards). *Chr10:104,845,940-104,953,056* (GRCh37/hg19 by Ensembl<sup>117</sup>). No known monogenic conditions reported in OMIM (OMIM). *NT5C2* is ubiquitously expressed (GTEx) with the highest expression observed in the thyroid and esophagus. Mouse knockout models demonstrate reduced body weight gain, insulin resistance on high-fat diet, and white adipose tissue mass<sup>78,97</sup>. **In-vitro studies in human skeletal muscle tissue show a suppression of 5'-Nucleotidase enzymes that promote AMP-activated protein kinase (AMPK) phosphorylation and metabolism<sup>118</sup>, which may suggest metabolic flexibility in an obese state.** Genetic variations in *NT5C2* have been associated with lower visceral and subcutaneous fat<sup>95</sup>, obesity, and the concurrence of obesity and depression<sup>96</sup>. Two GWAS in East Asian populations identified *NT5C2*, rs113278154, as “associated with metabolically unhealthy phenotypes among normal weight individuals”<sup>119,120</sup>. Rs11191548 of *NT5C2*, which is in complete LD (All pop:  $R^2=0.9815$ ) with the GWAS index SNP in the region, rs11191560, was studied to determine if miRNAs in the region disrupted binding to their target gene in an allele specific manner<sup>98</sup>. The study found that rs11191548 altered luciferase activity and decreased miRNA binding efficiency, and thus could explain how *NT5C2* may be one of the functional genes influencing BMI.

***YPEL3. Yippee-Like 3*** is a protein-coding gene that is involved in the proliferation and apoptosis in myeloid precursor cells (Genecards). It is required for central and peripheral glial cell development, and mutation of *YPEL3* causes neuropathy<sup>121</sup>. *chr16:30,103,635-30,108,236* (GRCh37/hg19 by Ensembl<sup>117</sup>). No known monogenic conditions reported in OMIM (OMIM). *YPEL3* is ubiquitously expressed (GTEx), with the highest expression observed in the whole blood and brain. Mouse knockout has a small body size and has neuronal irregularities, according to the International Mouse Phenotyping Consortium (IMPC). Liu and colleagues in 2020 identified *YPEL3* as a pleiotropic gene jointly influencing BMI and risk of schizophrenia, further supporting a neuronal correlation of this gene for obesity<sup>80</sup>. *YPEL3* knockdown in *Drosophila melanogaster* resulted in significant changes in body fat percentage<sup>79</sup>.

***ZNF646. Zinc Finger Protein 646*** is a protein-coding gene predicted to enable DNA-binding transcription factor activity, RNA polymerase II-specific, and RNA polymerase II cis-regulatory region sequence-specific DNA binding activity. *chr16:31,085,743-31,095,517* (GRCh37/hg19 by Ensembl). *ZNF646* is ubiquitously expressed (GTEx) with the highest expression observed in the testis. No known monogenic conditions reported in OMIM (OMIM). Expression of *ZNF646* has been associated with Parkinson’s disease in two different studies<sup>122,123</sup>.

***TMEM245. Transmembrane protein 245*** is a protein-coding gene with no known function. *chr9:111,777,432-111,882,225* (GRCh37/hg19 by Ensembl<sup>117</sup>). *TMEM245* is ubiquitously expressed (GTEx) with the highest expression in the thyroid, ovaries, and adrenal glands. Zhang and colleagues reported an association between *TMEM245* gene expression levels and atrial fibrillation<sup>101</sup>. Variants in this gene have been associated with schizophrenia<sup>102</sup>, age of menarche<sup>124</sup>, body height<sup>77</sup>, and cognitive abilities<sup>125</sup>.

***SPNS1. Lysosomal H(+)-carbohydrate transporter*** is a protein-coding gene that functions in lysosomal recycling at a late stage of autophagy (Genecards). *SPNS1* also functions as a sphingolipid transporter and may be involved in necrotic or autophagic cell death (Genecards). *Chr16:28,985,542-28,995,869* (GRCh37/hg19 by Ensembl<sup>117</sup>). No known monogenic conditions reported in OMIM (OMIM). *SPNS1* is ubiquitously expressed, with the strongest gene expression in arteries and the uterus. Differential expression of this gene has been associated with BMI<sup>126</sup>. This region has multiple coordinately regulated genes based on eQTL. In a recent in-silico study, *SPNS1* was differentially expressed in persons with T2D and obesity<sup>127</sup>. Polymorphisms in *SPNS1* have been associated with BMI<sup>128</sup>, asthma<sup>128</sup>, allergic disease<sup>129</sup>, and ADHD<sup>130</sup> in GWAS.

***GSTM3. Glutathione S-Transferase Mu 3*** is a protein coding gene, an enzyme that belongs to the mu class and functions in the detoxification of electrophilic compounds, including carcinogens, therapeutic drugs, environmental

toxins, and products of oxidative stress, by conjugation with glutathione. The genes encoding the mu class of enzymes are organized in a gene cluster on chromosome 1p13.3 and are known to be highly polymorphic. *GSTM3* may be involved in the uptake and detoxification of harmful compounds in the body at the testis and blood-brain barrier (Genecards). *chr1:110,276,554-110,284,384* (GRCh37/hg19 by Ensembl<sup>117</sup>). No known monogenic conditions reported in OMIM (OMIM). *GSTM3* is ubiquitously expressed with the strongest expression in the testis and ovaries. *GSTM3* has been associated with hyperinsulinemia, T2D<sup>131</sup>, and hypertension<sup>132</sup>. Recent studies have reported an increase of *GSTM3* in the omental fat of polycystic ovary syndrome (PCOS)<sup>133</sup>.

***SNAPC3. Small nuclear RNA activating complex polypeptide 3*** is part of the SNAPc complex required for the transcription of both RNA polymerase II and III small-nuclear RNA genes (Genecards). *SNAPC3* binds to the proximal sequence element (PSE), a non-TATA-box basal promoter element common to these 2 types of genes (GeneCard). *Chr9:15,422,702-15,465,951* (GRCh37/hg19 by Ensembl<sup>117</sup>). *SNAPC3* is ubiquitously expressed with the strongest expression in the testis and the cerebellum. Variants in this gene have been associated with schizophrenia<sup>99</sup>. DNA methylation in *SNAPC3* mediates the association between breastfeeding and early-life growth trajectories<sup>100</sup>.

## TABLES

**Table 1.** Summary table of significant genes and associations identified in discovery analysis and corresponding validation and generalization results. \*Filtering criteria for all validation and generalization analyses:  $P_{META} < 0.05$  &  $P_{META} < P_{SNP}$  &  $P_{META} < P_{BMI}$ .

Gene Symbol	GWAS SNP of locus	Number of SNPs interrogated in FHS	Number of significant SNP-transcript-BMI associations*				
			Discovery on FHS	Validation on CCHC	Generalization on liver	Generalization on brain – Hypothalamus	Generalization on brain – Nucleus Accumbens
<i>NT5C2</i>	rs11191560	115	115		103		
<i>GSTM3</i>	rs17024393	13	4				
<i>SNAPC3</i>	rs4740619	225	41	37	40		
<i>SPNS1</i>	rs3888190	91	91				
<i>TMEM245</i>	rs6477694	1	1		1		
<i>YPEL3</i>	rs4787491	10	10	10	10		10
<i>ZNF646</i>	rs9925964	46	46		15		

**Table 2.** Summary results for most significant SNP - expression - BMI combination identified in Discovery (FHS) sample. EA – effect allele, OA – other allele.

Gene	CHR	POS (b38)	dbSNP ID	EA	OA	GWAS tag SNP	GWAS tag dbSNP ID	Study	P <sub>BMI</sub>	N <sub>BMI</sub>	BETA <sub>BMI</sub>	P <sub>SNP</sub>	N <sub>SNP</sub>	BETA <sub>SNP</sub>	P <sub>META</sub>
<i>GSTM3</i>	chr1	109612066	rs17024393	C	T	Yes	rs17024393	FHS - Blood	2.67E-05	5,619	0.003	9.30E-13	5,257	0.123	<b>3.89E-15</b>
								CCHC - Blood	4.37E-01	934	-0.366	5.56E-03	769	0.757	1.70E-02
								Hypothalamus	9.48E-01	131	-7.858	-	-	-	-
								Nucleus Accumbens	1.61E-01	198	-7.033	-	-	-	-
								Liver	7.06E-01	2,225	0.007	6.56E-30		0.914	3.17E-28
<i>NT5C2</i>	chr10	103154183	rs74233809	C	T	No	rs11191560	FHS - Blood	3.48E-16	5,619	-0.004	5.82E-23	5,257	0.072	<b>1.16E-37</b>
								CCHC - Blood	1.07E-01	934	-0.803	7.98E-49	769	0.363	9.76E-48
								Hypothalamus	2.02E-01	131	-7.037	4.30E-01	108	0.128	2.99E-01
								Nucleus Accumbens	3.66E-01	198	-7.612	7.69E-04	170	0.276	2.59E-03
								Liver	6.40E-06	2,225	-0.044	4.87E-02	2,225	0.035	<b>4.98E-06</b>
<i>SNAPC3</i>	chr9	15786904	rs10962158	A	T	No	rs4740619	FHS - Blood	4.71E-10	5,619	0.002	2.97E-06	5,257	0.015	<b>3.66E-14</b>
								CCHC - Blood	1.63E-02	934	-0.559	2.96E-03	769	0.091	<b>5.28E-04</b>
								Hypothalamus	6.41E-01	131	-7.751	2.88E-01	108	0.083	4.97E-01
								Nucleus Accumbens	7.70E-01	198	-7.979	8.19E-01	170	0.011	9.21E-01
								Liver	4.03E-06	2,225	-0.033	3.79E-02	2,225	0.025	<b>2.55E-06</b>
<i>SPNS1</i>	chr16	28814099	rs11860513	T	C	No	rs3888190	FHS - Blood	1.33E-05	5,619	-0.001	1.03E-15	5,257	0.019	<b>4.59E-18</b>
								CCHC - Blood	4.12E-02	934	-1.266	4.98E-38	769	0.293	1.85E-37
								Hypothalamus	6.62E-01	131	-7.764	2.38E-01	108	0.119	4.49E-01
								Nucleus Accumbens	6.24E-01	198	-7.902	6.37E-01	170	-0.038	7.64E-01
								Liver	6.17E-01	2,225	0.006	-	-	-	-
<i>TMEM245</i>	chr9	109170062	rs6477694	T	C	Yes	rs6477694	FHS - Blood	2.39E-23	5,619	0.005	2.95E-36	5,257	0.055	<b>1.09E-57</b>
								CCHC - Blood	3.91E-02	934	0.396	2.17E-20	769	0.337	4.20E-20

								Hypothalamus	7.22E-01	131	-7.796	5.13E-01	108	0.059	7.38E-01
								Nucleus Accumbens	2.88E-01	198	-7.455	1.89E-01	170	0.062	2.13E-01
								Liver	<b>1.72E-07</b>	2,225	-0.341	3.78E-34	2,225	0.113	<b>6.08E-39</b>
<i>YPEL3</i>	chr16	29986879	rs4077410	G	A	No	rs4787491	FHS - Blood	3.19E-25	5,619	0.007	1.68E-21	5,257	0.057	<b>5.98E-46</b>
								CCHC - Blood	1.07E-01	934	-0.803	3.21E-02	769	0.037	<b>2.30E-02</b>
								Hypothalamus	3.22E-01	131	-7.365	1.07E-01	108	0.144	1.50E-01
								Nucleus Accumbens	1.04E-01	198	-6.692	1.23E-02	170	0.156	<b>9.79E-03</b>
								Liver	<b>1.09E-08</b>	2,225	-0.127	1.39E-52	2,225	0.163	<b>2.10E-58</b>
<i>ZNF646</i>	chr16	31077026	rs749671	A	G	No	rs9925964	FHS - Blood	9.68E-20	5,619	-0.003	1.89E-09	5,257	0.017	<b>7.33E-26</b>
								CCHC - Blood	3.91E-02	934	0.396	6.16E-01	769	-0.010	1.14E-01
								Hypothalamus	5.83E-01	131	-7.708	4.44E-01	108	-0.042	6.09E-01
								Nucleus Accumbens	2.72E-01	198	-7.414	7.26E-02	170	-0.067	9.71E-02
								Liver	9.25E-03	2,225	-0.020	7.32E-50	2,225	-0.195	8.05E-50